

Bias associated with delayed verification in test accuracy studies: accuracy of tests for endometrial hyperplasia may be much higher than we think!

Clark, Thomas; ter Riet, Gerben; Coomarasamy, Arri; Khan, Khalid

DOI:

[10.1186/1741-7015-2-18](https://doi.org/10.1186/1741-7015-2-18)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Clark, T, ter Riet, G, Coomarasamy, A & Khan, K 2004, 'Bias associated with delayed verification in test accuracy studies: accuracy of tests for endometrial hyperplasia may be much higher than we think!', *BMC medicine*, vol. 2, 18. <https://doi.org/10.1186/1741-7015-2-18>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

© 2004 Clark et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Checked July 2015

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Research article

Open Access

Bias associated with delayed verification in test accuracy studies: accuracy of tests for endometrial hyperplasia may be much higher than we think!

T Justin Clark^{*1}, Gerben ter Riet², Aravinthan Coomarasamy¹ and Khalid S Khan¹

Address: ¹Academic Department of Obstetrics & Gynaecology, Birmingham Women's Hospital, Birmingham, B15 2TG, UK and ²Academic Medical Center, Department of General Practice, 1105 AZ, Amsterdam, Netherlands

Email: T Justin Clark^{*} - justin.clark@bwhct.nhs.uk; Gerben ter Riet - G.terRiet@amc.uva.nl; Aravinthan Coomarasamy - a.coomarasamy@bham.ac.uk; Khalid S Khan - khalid.khan@bwhct.nhs.uk

^{*} Corresponding author

Published: 11 May 2004

BMC Medicine 2004, 2:18

Received: 16 September 2003

Accepted: 11 May 2004

This article is available from: <http://www.biomedcentral.com/1741-7015/2/18>

© 2004 Clark et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: To empirically evaluate bias in estimation of accuracy associated with delay in verification of diagnosis among studies evaluating tests for predicting endometrial hyperplasia.

Methods: Systematic reviews of all published research on accuracy of miniature endometrial biopsy and endometrial ultrasonography for diagnosing endometrial hyperplasia identified 27 test accuracy studies (2,982 subjects). Of these, 16 had immediate histological verification of diagnosis while 11 had verification delayed > 24 hrs after testing. The effect of delay in verification of diagnosis on estimates of accuracy was evaluated using meta-regression with diagnostic odds ratio (dOR) as the accuracy measure. This analysis was adjusted for study quality and type of test (miniature endometrial biopsy or endometrial ultrasound).

Results: Compared to studies with immediate verification of diagnosis (dOR 67.2, 95% CI 21.7–208.8), those with delayed verification (dOR 16.2, 95% CI 8.6–30.5) underestimated the diagnostic accuracy by 74% (95% CI 7%–99%; *P* value = 0.048).

Conclusion: Among studies of miniature endometrial biopsy and endometrial ultrasound, diagnostic accuracy is considerably underestimated if there is a delay in histological verification of diagnosis.

Background

The natural history of endometrial hyperplasia is not fully understood [1]. What is known is that a proportion of simple and complex hyperplastic processes will regress without treatment [2] although the time scale over which such regression may occur is unclear. Similarly the time scale over which benign endometrium progresses to hyperplasia is also unknown. Among studies evaluating

accuracy of tests for diagnosis of hyperplasia (miniature biopsy or ultrasonography), it has previously been hypothesised that if histological verification of diagnosis after performing the test is delayed, the estimation of test accuracy may be influenced by the phenomena of disease regression or progression [3]. For instance, false positive diagnoses of endometrial hyperplasia may occur due to natural disease regression during the time interval

between testing and verification of diagnosis. Similarly, false negative diagnoses may also result from progression of benign functional or atrophic endometrium.

To obtain accurate estimates of test accuracy in studies of hyperplasia, an immediate comparison of the test under scrutiny with a reference standard that verifies the diagnosis will be essential [4-6]. When accuracy studies suffer from a delay in performance of the reference standard, the resultant false positives and false negatives will be expected to lead to an underestimation of test accuracy. In systematic reviews, when studies of various designs are collated, the extent of underestimation that arises from delay is important in obtaining an unbiased pooled accuracy estimate. To our knowledge, the extent of underestimation of accuracy due to a delay in verification of diagnosis has not been evaluated empirically in studies of endometrial hyperplasia. We undertook this analysis to examine formally how inaccurate the estimation of accuracy can be in studies evaluating miniature endometrial biopsy devices and endometrial thickness measurement by pelvic ultrasonography for predicting endometrial hyperplasia when there are delays in histological verification of diagnosis.

Methods

To test our hypothesis, a data set of all the published studies reporting the accuracy of miniature endometrial biopsy devices and endometrial ultrasonography for predicting endometrial hyperplasia was obtained from systematic reviews [7,8]. The reviews focused on test accuracy studies in which the results of the test were compared with the results of a reference standard. The targeted population was women with abnormal pre- or postmenopausal uterine bleeding. The diagnostic tests of interest were miniature endometrial biopsy devices (for example, pipelle® endometrial suction curette, Unimar, Wilton, CT, USA) and endometrial thickness measurement by pelvic ultrasonography. The reference standard was endometrial histology obtained by an independent endometrial sampling technique, for example, inpatient curettage (with hysteroscopy) or hysterectomy.

Identification of studies

Two independent electronic searches of MEDLINE and EMBASE were conducted to identify relevant citations on endometrial biopsy (1980–1999) and ultrasonography (1966–2000). Search term combination for endometrial biopsy [8] was *diagnosis* (MeSH) AND *endometrial biopsy* (textword), while that for studies on ultrasonography [7] was *ultrasound* AND *endometrial thickness* AND *sonography* (textwords). The searches were limited to human studies, but there were no language restrictions. Relevant studies were identified by examining all the retrieved citations, reference lists of all known reviews and primary studies,

and direct contact with manufacturers. Details of the search and selection processes can be found in the published reports of the reviews [7,8].

Study quality assessment

All selected studies were assessed for their methodological quality defined as the confidence that study design, conduct and analysis minimize bias in the estimation of diagnostic accuracy [9-11]. We considered the following features in quality assessment: method of recruitment of sample, appropriateness of patient spectrum, and blinding of comparison between test and reference standard. Recruitment was considered to be adequate if patient selection was consecutive or a random sample was obtained. Patient spectrum was considered to be appropriate if both pre- and postmenopausal women were included. Blinding was considered to be present if it was clearly reported that the pathologists providing histological reports were kept unaware of the results of miniature endometrial biopsy or endometrial ultrasonography. If the results of the diagnostic tests were divulged to the pathologists or in the absence of any such reporting, blinding was categorised as absent. For the purpose of our analysis, studies were classified into two quality categories: Category I studies had any one of the following features: adequate recruitment, appropriate spectrum, and blinding; category II studies had none of the above quality features.

Data extraction

In addition to assessment of methodological quality, data were extracted to allow classification of studies into one of two groups: i) immediate verification – reference standard performed within 24 hours of testing, and ii) delayed verification – reference standard performed more than 24 hours after testing. Any studies that could not be categorised in this way due to lack of reporting were excluded. Data were then abstracted as 2×2 tables and estimates of diagnostic accuracy were derived for each individual study. A correction factor of 0.5 was used when cells of the 2×2 tables included zero values [12]. True positive rates (sensitivity), false positive rates (1-specificity) and diagnostic odds ratios (dORs) were calculated for each primary evaluation. The dOR represents a ratio of the positive and negative likelihood ratios and it can be mathematically summarised as:

$$\text{dOR} = [\text{sensitivity}/(1-\text{specificity})] / [(1-\text{sensitivity})/\text{specificity}]$$

Statistical analysis

Pooled dORs were generated as the principal measures of diagnostic accuracy. Meta-analyses to produce summary estimates of accuracy were performed separately for subgroups of studies reporting immediate and delayed

verification. To delineate the impact of delay in verification of diagnosis, we used meta-regression analysis [13,14] with the log of dOR as the accuracy measure. This technique fitted a multivariable linear regression model for examining the influence of delay, quality and test type on the estimation of accuracy observed among studies included in the analysis (random effects model). In this way the analysis was adjusted for the confounding effects of study quality (two quality categories) and type of test (miniature endometrial biopsy or endometrial ultrasound).

Results

Selection of studies

The study selection process is shown in Figure 1. In total there were 2,982 subjects in 27 diagnostic evaluations reported in the 24 eligible primary studies. Eleven evaluations delayed verification of the diagnosis by more than 24 hours; the delay was up to six months in one study, up to four weeks in four studies, up to three weeks in one study and up to one week in the remaining three studies. Three of these studies were rated as category I for methodological quality, and eight as category II. Sixteen evaluations verified the diagnosis within 24 hours of the test. Among these, seven studies were rated as category I for quality, and nine as category II (Table 1).

Table 2 shows the diagnostic accuracy results for individual studies according to test type and verification status in terms of delay. The summary statistics for the various subgroups showed that the dOR for studies with immediate verification was 67.2 (21.7–208.8) while that for studies with delayed verification was 16.2 (8.6–30.5) as shown in Figure 2. Meta-regression analysis for bias due to delay in verification of diagnosis, adjusted for study quality and test type, showed that the underestimation of test accuracy among studies with delayed verification was 74% (95% CI 7%-99%; $P = 0.048$) on average compared to studies with immediate verification (Table 3).

Discussion

Our study shows empirically the magnitude of bias associated with delay in verification of diagnosis in test accuracy studies. Delay in verification of more than 24 hours was associated with a considerable underestimation of accuracy of miniature biopsy and endometrial ultrasonography in diagnosing endometrial hyperplasia. This supports the premise that the reported limited accuracy of miniature endometrial biopsy devices and endometrial ultrasonography in diagnosing hyperplasia is due, in part, to natural history of disease rather than resulting entirely from intrinsic problems with performance of the diagnostic tools [3].

We posed our hypothesis *a priori* and tested it in as rigorous a manner as possible. Our literature search was without language restriction, facilitating retrieval of many relevant test accuracy studies. However, due to poverty of reporting many critical pieces of information were missing in the available literature, restricting the number of studies that could be included in our analysis (for example, 31 studies were ineligible for inclusion because explicit information about time before verification was omitted). Our examination of delays in verification was also restricted; just two time categories were discernible (delay < 24 hours or > 24 hours). Immediate verification (reference standard to be performed straight after the index test) was not achievable in some studies because the reference test (inpatient endometrial sampling) necessitated use of general anaesthesia. A practical cut-off of 24 hours was taken to allow time for reference testing to be undertaken when the preceding index tests (miniature endometrial biopsy and ultrasound) were performed in the conscious outpatient. Although the natural history of endometrial hyperplasia is unclear, it is unlikely that biological alteration would have occurred within 24 hours. To study the rate of disease progression or regression would require repeated testing over time, but such a study is unlikely to be ethically justifiable, given that most clinicians will institute treatment following initial diagnosis. Such a study would then become one of prognosis under treatment rather than a natural history study.

We also evaluated other features of methodological quality and, in general, found the quality of studies to be poor. For example, only three studies reported blinding interpretation of the reference test from knowledge of results from the index test. A lack of blinding can introduce bias and overestimation of diagnostic accuracy [4]. Pathological interpretation of endometrial hyperplasia is open to a varying degree of subjectivity especially at extreme ends of the spectrum, where overlap with benign functional endometrium (simple hyperplasia) and cancer (complex hyperplasia with cytological atypia) is more likely. Absence (or explicit reporting) of blinding is thus associated with poorer methodological quality and this feature was incorporated in our quality assessment. Our analysis adjusted for the confounding effects of quality but our inferences should be interpreted with caution due to relative scarcity of good quality studies.

Conclusions

Our findings have implications for research into new diagnostic interventions. Our results demonstrate that test evaluation with robust study design (immediate verification) showed good test performance but evaluation in poor designs (delayed verification) showed poor performance. Poor designs may reflect the situation prevalent in routine clinical practice where test results may not be

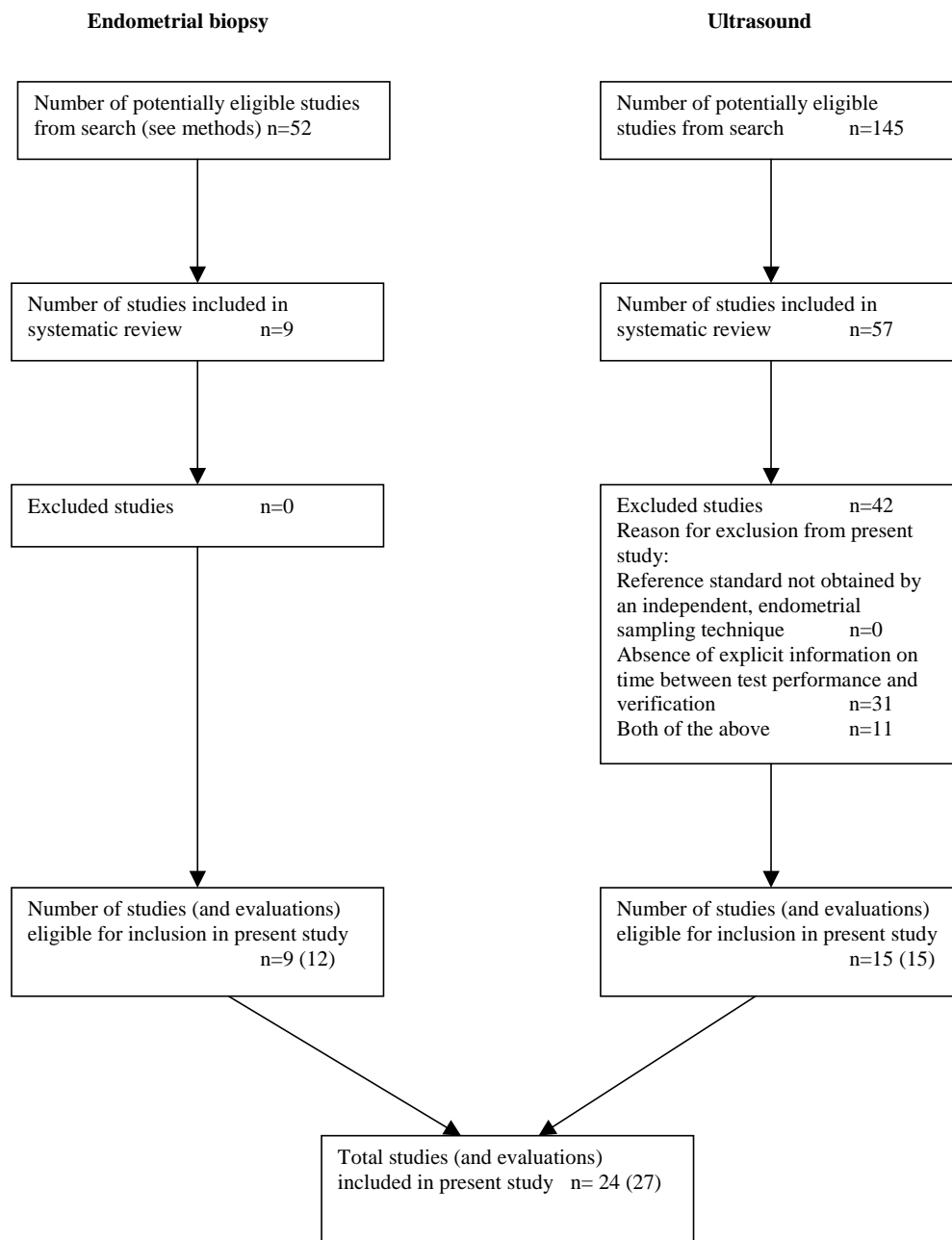


Figure 1
Flow diagram showing study selection process.

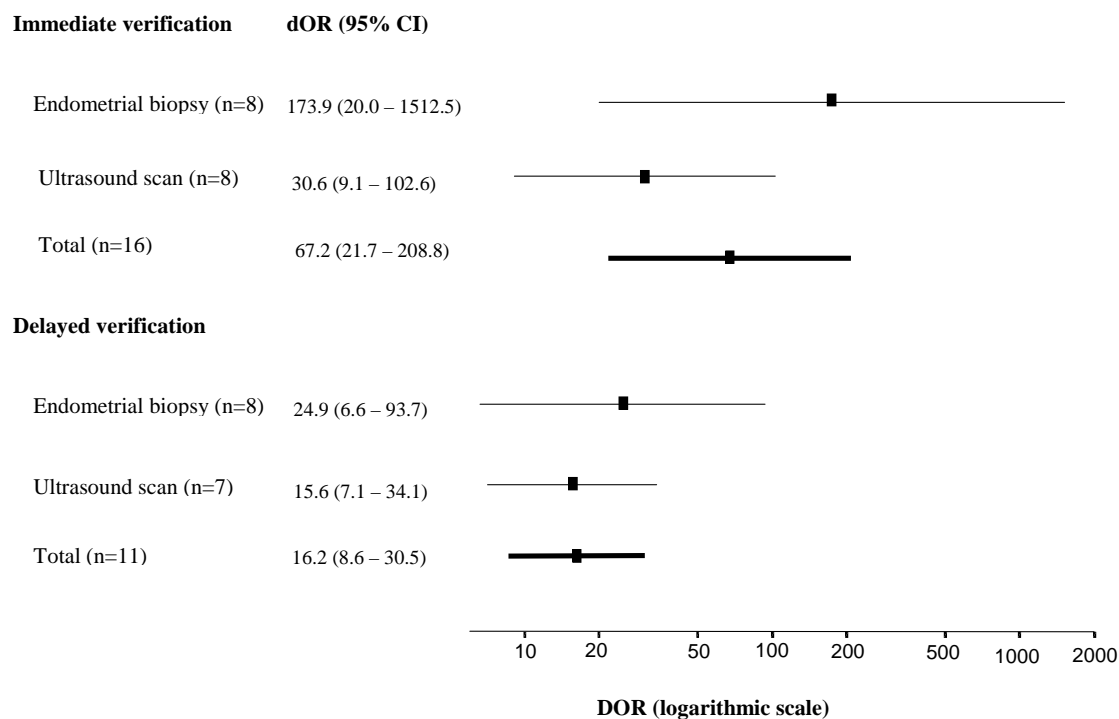
Table 1: Study characteristics and methodological quality.

Study (Year published)	Details	Delay (hours)	Patient selection	Bleeding type / Menopausal status (%)				Reference standard	Blinding of results	Study quality level
				Post	HRT	Pre	†Other			
IMMEDIATE VERIFICATION (</=24 HOURS) (16 studies)										
Endometrial biopsy studies (8)										
Sun-Kuie et al. [15] (1992)	Gynoscann®	Day before	Unreported	*5 (11)	-	41 (89)	-	D&C	Unreported	I
Goldberg et al. [16] (1981)	Accurette®	Immediate	Unreported	30 (100)	-		-	D&C	Unreported	II
Sonnendecker et al. [17] (1981)	Accurette®	Immediate	Unreported	*6 (24)	-	17 (76)	-	D&C	Unreported	I
Kufahl et al. [18] (1997)	Explora®	Immediate	Consecutive	*33 (21)	-	125 (79)	-	Hysterectomy	Unreported	I
Kufahl et al. [18] (1997)	Gynoscann®	Immediate	Consecutive	*36 (21)	-	133 (79)	-	Hysterectomy	Unreported	I
Goldschmit et al. [19] (1993)	Pipelle®	Immediate	Consecutive	*34 (23)	-	115 (77)	-	D&C	Yes	I
Kavak et al. [20] (1996)	Pipelle®	Immediate	Unreported	*34 (56)	-	27 (44)	-	D&C	Yes	I
Goldberg et al. [16] (1981)	Vabra Aspirator®	Immediate	Unreported	31 (100)	-	-	-	D&C	Unreported	II
Ultrasound scan studies (8)										
Botsis et al. [21] (1992)	≤ 4 mm DL	Day before	Unreported	120(100)	-	-	-	D&C	Unreported	II
Garuti et al. [22] (1999)	≤ 4 mm DL	Immediate	Unreported	368 (88)	51 (12)	-	-	D&C	Unreported	I
Haller et al. [23](1996)	≤ 4 mm DL	Day before	Unreported	81 (100)	-	-	-	D&C	Unreported	II
Grigoriou et al. [24] (1996)	≤ 5 mm DL	Day before	Unreported	250(100)	-	-	-	D&C	Unreported	II
Karlsson et al. [25](1993)	≤ 5 mm DL	Day before	Unreported	103(100)	-	-	-	D&C	Unreported	II
Malinova et al. [26] (1996)	≤ 5 mm DL	Day before	Unreported	154(100)	-	-	-	D&C	Unreported	II
Wolman et al. [27] (1996)	≤ 5 mm DL	Immediate	Unreported	54 (100)	-	-	-	D&C	Unreported	II
Malinova et al. [28] (1995)	≤ 5 mm SL	Day before	Unreported	118(100)	-	-	-	D&C	Unreported	II
DELAYED VERIFICATION (> 24 HOURS) (11 studies)										
Endometrial biopsy studies (4)										
Stovall et al. [29] (1989)	Novak Curette®	Unreported	Unreported	-	-	-	165(100) NS	Hysterectomy	Unreported	II
Krampl et al. [3] (1997)	Pipelle®	< 6 months	Unreported	37 (12)	-	247 (77)	35 (11)	TCRE/ Hysterectomy	Unreported	I
Gupta et al. [30] (1996)	Pipelle®	< 1 month	Unreported	54 (100)	-	-		D&C	Unreported	II
Stovall et al. [29] (1989)	Vabra Aspirator®	Unreported	Unreported	-	-	-	62 (100) NS	Hysterectomy	Unreported	II

Table 1: Study characteristics and methodological quality. (Continued)

Ultrasound scan studies (7)											
Guner <i>et al.</i> [31] (1996)	≤ 4 mm DL	≤ 7 days	Unreported	192 (100)	-	-	-	D&C	Unreported	II	
Abu-Ghazze <i>et al.</i> [32] (1999)	≤ 5 mm DL	1 week	Unreported	98 (100)	-	-	-	D&C	Unreported	II	
DeSilva <i>et al.</i> [33] (1997)	≤ 5 mm DL	< 4 weeks	Consecutive	44 (88)	6 (12)	-	-	D&C	Unreported	I	
Gupta <i>et al.</i> [30] (1996)	≤ 5 mm DL	< 1 month	Unreported	75 (100)	-	-	-	D&C	Yes	I	
Taviani <i>et al.</i> [34] (1995)	≤ 5 mm DL	1 week	Unreported	41 (100)	-	-	-	D&C	Unreported	II	
Moreles <i>et al.</i> [35] (1998)	≤ 6 mm DL	< 3 weeks	Unreported	200 (100)	-	-	-	D&C	Unreported	II	
Mortakis <i>et al.</i> [36] (1997)	≤ 3 mm SL	< 4 weeks	Unreported	78 (100)	-	-	-	D&C	Unreported	II	

*Numbers calculated from initial proportion of patients within these groups before missing outcome data was excluded. †Other refers to proportion of women included in the study who did not have abnormal uterine bleeding as an indication for investigation. HRT, hormone replacement therapy; NS, not specified (refers to proportion of women included in the study where the type of abnormal uterine bleeding was not specified); DB, directed biopsy; D&C, dilatation and curettage; TCRE, transcervical resection of the endometrium; DL, double layer endometrial thickness; SL, single layer endometrial thickness.

**Figure 2**

Effect of delayed verification on the diagnostic accuracy of miniature endometrial biopsy and transvaginal ultrasound in detecting endometrial hyperplasia. Pooled diagnostic odds ratios (dOR) for studies with immediate and delayed verification.

Table 2: Accuracy stratified by time delay between test performance and confirmation by chosen reference test histology.

Device (no. evaluations) & study (year published)	+ve test (sensitivity)	-ve test (1-specificity)	Odds ratio (95% CI)
IMMEDIATE VERIFICATION (<= 24 HOURS) (16 studies)			
Endometrial biopsy studies (8)			
Sun-Kuie et al. [15] (1992)	2/4 (0.5)	0/42 (0.0)	85.0 (3.2–2289.6)
Goldberg et al. [16] (1981)	5/5 (1.0)	0/25 (0.0)	561.0 (10.0–31463.1)
Sonnendecker et al. [17] (1981)	2/2 (1.0)	0/21 (0.0)	215.0 (3.5–13408.5)
Kufahl et al. [18] (1997)	14/15 (0.9)	4/143 (0.03)	486.5 (50.8–4658.5)
Kufahl et al. [18] (1997)	3/15 (0.2)	15/154 (0.1)	2.3 (0.6–9.1)
Goldschmit et al. [19] (1993)	11/14 (0.8)	1/135 (0.01)	491.3 (47.1–5127.3)
Kavak et al. [20] (1996)	4/4 (1.0)	0/57 (0.0)	1035.0 (18.3–58563.1)
Goldberg et al. [16] (1981)	6/6 (1.0)	0/25 (0.0)	663.0 (12.0–36690.0)
Ultrasound scan studies (8)			
Botsis et al. [21] (1992)	10/10 (1.0)	4/92 (0.04)	413.0 (20.8–8221.1)
Garuti et al. [22] (1999)	44/46 (0.96)	196/313 (0.63)	13.1 (3.1–55.2)
Haller et al. [23] (1996)	14/16 (0.88)	34/49 (0.69)	3.1 (0.6–15.3)
Grigoriou et al. [24] (1996)	45/45 (0.98)	30/181 (0.17)	452.0 (27.1–7538.5)
Karlsson et al. [25] (1993)	10/10 (1.0)	21/78 (0.27)	56.2 (3.2–1000.5)
Malinova et al. [26] (1996)	11/11 (1.0)	32/74 (0.43)	30.1 (1.7–529.5)
Wolman et al. [27] (1996)	10/11 (0.91)	12/39 (0.31)	22.5 (2.6–196.1)
Malinova et al. [28] (1995)	7/7 (1.0)	19/54 (0.35)	27.31 (1.5–504.1)
DELAYED VERIFICATION (> 24 HOURS) (11 studies)			
Endometrial biopsy studies (4)			
Stovall et al. [29] (1989)	8/16 (0.5)	4/149 (0.03)	36.3 (9.0–146.3)
Krampl et al. [3] (1997)	14/35 (0.4)	24/284 (0.0)	7.2 (3.3–16.0)
Gupta et al. [30] (1996)	6/10 (0.6)	1/44 (0.02)	64.5 (6.1–677.6)
Stovall et al. [29] (1989)	7/7 (1.0)	7/55 (0.13)	97.0 (5.0–1879.9)
Ultrasound scan studies (7)			
Guner et al. [31] (1996)	31/31 (1.0)	61/142 (0.43)	83.5 (5.0–1391.3)
Abu-Ghazze et al. [32] (1999)	2/2 (1.0)	58/95 (0.61)	3.2 (0.2–68.6)
DeSilva et al. [33] (1997)	2/3 (0.67)	10/44 (0.23)	6.8 (0.6–83.0)
Gupta et al. [30] (1996)	4/4 (1.0)	22/68 (0.32)	18.6 (1.0–360.7)
Taviani et al. [34] (1995)	2/2 (1.0)	16/37 (0.43)	6.5 (0.3–145.1)
Moreles et al. [35] (1998)	33/37 (0.9)	37/143 (0.26)	23.6 (7.8–71.2)
Mortakis et al. [36] (1997)	4/5 (0.8)	26/66 (0.39)	6.2 (0.7–58.2)

Table 3: Results of meta-regression analysis.

Outcome Explanatory variables	Univariable analysis		Multivariable analysis	
	Ratio dOR (95% CI) [†]	P value	Ratio dOR (95% CI) [†]	P value
Endometrial hyperplasia				
• Delay in verification of diagnosis (> 24 hours vs. < 24 hours)	0.31 (0.08–0.84)	0.089	0.26 (0.07–0.99)	0.048
• Study quality (category II vs. category I)	1.36 (0.34–5.53)	0.664	3.46 (0.79–15.0)	0.098
• Test (Endometrial biopsy vs. ultrasound endometrial thickness)	3.22 (0.83–12.55)	0.091	5.0 (1.2–20.7)	0.027

[†]The dependent variable is the natural logarithm of the diagnostic odds ratio (dOR). Results are presented as the ratio of diagnostic odds ratios (RdOR); an RdOR < 1 means that the diagnostic accuracy is reduced and a RdOR > 1 means that it is increased in relation to the reference category, < 24 h, category I, and ultrasound endometrial thickness, respectively.

immediately confirmed due to resource and other implications. Thus diagnostic evaluations carried out in routine practice may mask the accuracy of tests.

Competing interests
None declared.

Authors' contributions

TJC and KSK conceived and designed the study with input from GtR. TJC conducted the systematic review and acquired all data. AC conducted the statistical analyses with input from KSK and GtR. TJC wrote all versions of the manuscript. KSK and GtR critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

References

1. Fox H: **Endometrial hyperplasia: a conceptual and practical approach.** *Gynaecology Forum* 1996, **1**:7-9.
2. Kurman RJ, Kaminski PF, Norris HJ: **The behaviour of endometrial hyperplasia: a long-term study of "untreated" hyperplasia in 170 patients.** *Cancer* 1985, **56**:403-412.
3. Krampl E, Soby B, Istre O: **How representative are pipelle endometrial biopsies? A retrospective analysis of 324 biopsies followed by transcervical resection of the endometrium or hysterectomy.** *Gynaecol Endosc* 1997, **6**:277-281.
4. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, Van der Meulen JHP, Bossuyt PM: **Empirical evidence of design-related bias in studies of diagnostic tests.** *JAMA* 1999, **282**:1061-1066.
5. Reid MC, Lachs MS, Feinstein AR: **Use of methodological standards in diagnostic test research. Getting better but still not good.** *JAMA* 1995, **274**:645-651.
6. Sheps SB, Schechter MT: **The assessment of diagnostic tests: a survey of current medical research.** *JAMA* 1984, **252**:2418-2422.
7. Gupta JK, Chien PF, Voit D, Clark TJ, Khan KS: **Ultrasonographic endometrial thickness for diagnosing endometrial pathology in women with postmenopausal bleeding: a meta-analysis.** *Acta Obstet Gynecol Scand* 2002, **81**:799-816.
8. Clark TJ, Mann CH, Shah N, Song F, Khan KS, Gupta JK: **Accuracy of outpatient endometrial biopsy in the diagnosis of endometrial hyperplasia: a systematic quantitative review.** *Acta Obstet Gynecol Scand* 2001, **80**:784-793.
9. **Cochrane Methods Working Group on Systematic Reviews of Screening and Diagnostic Tests: Recommended Methods** [<http://www.cochrane.org/cochrane/sadt/doc1.htm>]
10. Dunn G, Everitt B: *Clinical Biostatistics.* London: Edward Arnold 1995.
11. Khan KS, ter Riet G, Popay J, Nixon J, Kleijnen J: **Study quality assessment.** In *Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews* 2nd edition. Edited by: Khan KS, Ter Riet G, Glanville J, Sowden AJ, Kleijnen J. York: NHS Centre for Reviews and Dissemination (CRD), University of York; 2001.
12. Sankey S, Weisfields L, Fine M, Kapoor W: **An assessment of the use of the continuity correction for sparse data in meta analysis.** *Comm Stat Sim Comput* 1996, **25**:1031-1056.
13. Sterne JAC, Egger M, Davey Smith G: **Investigating and dealing with publication and other biases.** In *Systematic Reviews in Health Care: Meta-analysis in Context* Edited by: Egger M, Davey Smith G, Altman DG. London: BMJ Publishing Group; 2001:189-208.
14. Song F, Sheldon T, Sutton A, Abrams K, Jones D: **Methods for exploring heterogeneity in meta-analysis.** *Eval Health Prof* 2001, **24**:126-151.
15. Sun-Kuie T, Sian-Ann T, Ka-Mui C, Soo-Kim L: **The diagnostic value and patient acceptability of outpatient endometrial sampling with Gynoscann.** *Aust NZ J Obstet Gynaecol* 1992, **32**:73-76.
16. Goldberg GL, Tsalacopoulos GDDA: **A comparison of the Accurette and Vabra aspirator and uterine curettage.** *S Afr Med J* 1982, **61**:114-116.
17. Sonnendecker EW, Sevit H, Hofmeyr GJ: **Diagnostic accuracy of the accurette endometrial sampler.** *S Afr Med J* 1982, **61**:109-113.
18. Kufahl J, Pederson I, Eriksen PS, Helkjaer PE, Larsen LG, Jensen KL, de Nully P, Philipson T, Wahlin A: **Transvaginal ultrasound, endometrial cytology sampled by Gynoscann and histology obtained by Uterine Explora Curette compared to the histology of the uterine specimen. A prospective study in pre- and postmenopausal women undergoing elective hysterectomy.** *Acta Obstet Gynecol Scand* 1997, **76**:790-796.
19. Goldchmit R, Katz Z, Blickstein I, Caspi B, Dgani R: **The accuracy of endometrial Pipelle sampling with and without sonographic measurement of endometrial thickness.** *Obstet Gynecol* 1993, **82**:727-730.
20. Kavak Z, Ceyhan N, Pekin S: **Combination of vaginal ultrasonography and pipelle sampling in the diagnosis of endometrial disease.** *Aust N Z J Obstet Gynaecol* 1996, **36**:63-66.
21. Botsis D, Kassanos D, Pyrgiotis E, Zourlas PA: **Vaginal sonography of the endometrium in postmenopausal women.** *Clin Exp Obstet Gynecol* 1992, **19**:189-192.
22. Garuti G, Sambruni I, Cellani F, Garzia D, Alleva P, Luerti M: **Hysteroscopy and transvaginal ultrasonography in postmenopausal women with uterine bleeding.** *Int J Gynaecol Obstet* 1999, **65**:25-33.
23. Haller H, Matejic N, Rukavina B, Krasic M, Rupcic S, Mozetic D: **Transvaginal sonography and hysteroscopy in women with postmenopausal bleeding.** *Int J Gynecol Obstet* 1996, **54**:155-159.
24. Grigoriou O, Kalovidouras A, Papadias C, Antoniou G, Antonaki V, Giannikos L: **Transvaginal sonography of the endometrium in women with postmenopausal bleeding.** *Maturitas* 1996, **23**:9-14.
25. Karlsson B, Granberg S, Wiklund M, Ryd W, Norstrom A: **Endovaginal scanning of the endometrium compared to cytology and histology in women with postmenopausal bleeding.** *Gynecol Oncol* 1993, **50**:173-178.
26. Malinova M, Pehlivanov B: **Transvaginal sonography and progesterone challenge for identifying endometrial pathology in postmenopausal women.** *Int J Gynecol Obstet* 1996, **52**:49-53.
27. Wolman I, Sagi J, Ginat S, Jaffa AJ, Hartoov J, Jedwab G: **The sensitivity and specificity of vaginal sonography in detecting endometrial abnormalities in women with postmenopausal bleeding.** *J Clin Ultrasound* 1996, **24**:79-82.
28. Malinova M, Pehlivanov B: **Transvaginal sonography and endometrial thickness in patients with postmenopausal uterine bleeding.** *Eur J Obstet Gynecol Reprod Biol* 1995, **58**:161-165.
29. Stovall TG, Solomon SK, Ling FW: **Endometrial sampling prior to hysterectomy.** *Obstet Gynecol* 1989, **73**:405-409.
30. Gupta JK, Wilson S, Desai P, Hau C: **How should we investigate women with postmenopausal bleeding?** *Acta Obstet Gynecol Scand* 1996, **75**:475-479.
31. Guner H, Tiras MB, Karabacak O, Sarikaya H, Erdem M, Yildirim M: **Endometrial assessment by vaginal ultrasonography might reduce endometrial sampling in patients with postmenopausal bleeding: A prospective study.** *Aust N Z J Obstet Gynaecol* 1996, **36**:175-178.
32. Abu-Ghazze Y, Shakoury WA, Barqawi R: **Comparative study of transvaginal hysterosonography and biopsy for the evaluation of post-menopausal bleeding.** *Ann Saudi Med* 1999, **19**:116-119.
33. De Silva BY, Stewart K, Steven JD, Sathanandan M: **Transvaginal ultrasound measurement of endometrial thickness and endometrial pipelle sampling as an alternative diagnostic procedure to hysteroscopy and dilatation and curettage in the management of post-menopausal bleeding.** *J Obstet Gynaecol* 1997, **17**:399-402.
34. Taviani A, Braccini S, Toniazzi P, Pantani P, Costamagna V, Gambini G, Pancanti V: **Transvaginal echography in patients with postmenopausal metrorrhagia.** *Minerva Ginecol* 1995, **47**:369-372.
35. Morales FJ, Dualde D, Marinero A: **Value of vaginal ultrasound in the diagnosis of postmenopausal metrorrhagia.** *Radiologia* 1998, **40**:255-262.
36. Mortakis AE, Mavrelis K: **Transvaginal ultrasonography and hysteroscopy in the diagnosis of endometrial abnormalities.** *J Am Assoc Gynecol Laparosc* 1997, **4**:449-452.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1741-7015/2/18/prepub>